指数损失下总体的最优估计

Yimeng Ren

1 指数损失下总体的最优估计

损失函数：$L(y, f(x)) = \exp[-yf(x)] \quad y \in \{\pm 1\}$

相当于回归中的残差损失（?）

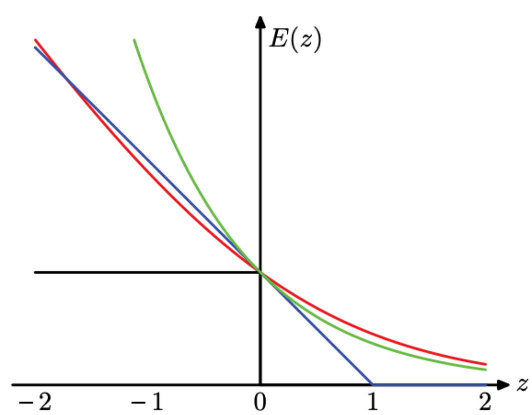

Figure 3: Loss functions for learning: Black: 0-1 loss. Blue: Hinge Loss. Red: Logistic regression. Green: Exponential loss. (Figure from *Pattern Recognition and Machine Learning* by Chris Bishop.)

图 1: 二分类问题的几种损失函数

结论：

$$f^*(x) = argmin_{f(x)}[E_{Y|X}(e^{-yf(x)})] = \frac{1}{2}\log\frac{p(y=1|x)}{p(y=-1|x)}$$

$$\Leftrightarrow p(y = 1 \mid x) = \frac{1}{1 + e^{-2f^*(x)}}$$

下证：

$$argmin_{f(x)}[E_{Y|X}(e^{-yf(x)})] = \frac{1}{2}\log\frac{p(y=1|x)}{p(y=-1|x)}$$

$$E_{Y|X}\left(e^{-yf(x)}\right) = P(y=1 \mid x)e^{-f(x)} + p(y=-1 \mid x)\cdot e^{f(x)}$$

$$\frac{\partial E_{Y|X}(e^{-yf(x)})}{\partial f(x)} = -p(y=1 \mid x)\cdot e^{-f(x)} + p(y=-1 \mid x)e^{f(x)} = 0$$

$$\begin{aligned} &\Rightarrow \quad P(y=1 \mid x)e^{-f(x)} = P\left(y=-1 \mid x\right)e^{f(x)} \\ &\Rightarrow f^*(x) = \tfrac{1}{2}\log\tfrac{P(y=1|x)}{p(y=-1|x)} \end{aligned}$$

与 logistic regression 的关系:

logistic regression 中，$p = P(y=1 \mid x)$

$$\log\left(\frac{p}{1-p}\right) = f(x)$$

$$p = \frac{e^{f(x)}}{1+e^{f(x)}}, 1-p = \frac{1}{1+e^{f(x)}}$$

似然函数为:

$$L = p^y \cdot (1-p)^{1-y}, \quad y \in \{0,1\}$$

对数似然函数为:

$$\log(L) = y\log(p) + (1-y)\log(1-p)$$

注意，这里需要有 $y \in \{0,1\}$ 到 $y \in \{+1,-1\}$ 的转换如下:

$$if\ y' = 2y-1, then\ y' \in \{\pm 1\} \Leftrightarrow y = \frac{1}{2}\left(y'+1\right)$$

对数似然函数等价于:

$$\begin{aligned} &\log(L) = \tfrac{1}{2}\left(y'+1\right)\log(p) + \left[1-\tfrac{1}{2}\left(y'+1\right)\right]\log(1-p) \\ &= \left(\tfrac{y'+1}{2}\right)\log\left(\tfrac{1}{1+e^{-f(x)}}\right) + \left(\tfrac{1-y'}{2}\right)\cdot\log\left(\tfrac{1}{1+e^{f(x)}}\right) \\ &= \log\left(1+e^{-f(x)}\right)^{-\frac{y'+1}{2}} + \log\left(1+e^{f(x)}\right)^{-\frac{1-y'}{2}} \\ &= \begin{cases} \log\left(1+e^{-f(x)}\right)^{-1}, \quad y'=1 \\ \log\left(1+e^{f(x)}\right)^{-1}, \quad y'=-1 \\ = \log\left(1+e^{-y'f(x)}\right)^{-1} \end{cases} \end{aligned} \tag{1}$$

即 $\log(L)=\log\left(1+e^{-y'f(x)}\right)^{-1}$，最小化它和（Adaboost）指数损失下总体的最优估计是等价的。